



A hierarchical Bayesian approach to robust parameter design

Yuri Goegebeur, Peter Goos and Martina Vandebroek

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

A Hierarchical Bayesian Approach to Robust Parameter Design

Yuri Goegebeur^{*}, Peter Goos[†], Martina Vandebroek[‡]

June 1, 2007

Abstract

The goal of robust parameter design experiments is to identify significant location and dispersion factors that can be used to set the mean response at the target level and to decrease the sensitivity of the response to uncontrolled noise factors. We present a hierarchical Bayesian model and use empirical Bayes priors to find the active factors and to get reliable estimates of the location and dispersion parameters. The approach is particularly useful when the design points are not replicated, a case which is challenging with standard procedures.

Keywords: Hierarchical Bayesian model, Empirical Bayes priors, location and dispersion factors, Taguchi experiments, WinBUGS, Gibbs sampling, credible or posterior probability intervals

^{*}University of Southern Denmark, Katholieke Universiteit Leuven, Belgium

[†]Faculty of Applied Economics, Universiteit Antwerpen, Belgium

[‡]Faculty of Economics and Applied Economics, University Center for Statistics, Katholieke Universiteit Leuven, Belgium (martina.vandebroek@econ.kuleuven.be)

1 Introduction

The detection of active dispersion factors in addition to the identification of active location factors has received a lot of attention since Taguchi (1987) introduced the concept of robust parameter design. The key idea is that the performance of products and processes can be made robust or insensitive to the variation of noise factors, i.e. the factors that are hard to control and therefore vary randomly in practice. This can be achieved by an appropriate selection of the levels of the control or design factors which, unlike the noise factors, can be controlled during the experiment as well as outside the experimental context. Control factors that affect the mean performance of the product or process under investigation are then referred to as active or significant location factors, whereas control factors that have an impact on the variance are called active or significant dispersion factors.

The experimental design and data analysis methodology proposed by Taguchi (involving so-called crossed arrays or product designs, and signal-to-noise ratios) have been questioned frequently, see Nair (1992), Welch, Yu, Kang and Sacks (1990) and Shoemaker, Tsui and Wu (1991). Nevertheless, there is great consensus that the robust design philosophy is of utmost importance and robust parameter design experiments remain popular in industry.

Several authors have therefore presented alternative approaches for setting up and analyzing robust design experiments. Vining and Myers (1990) for instance propose a dual response approach in which a location model and a dispersion model are estimated separately. The dispersion model is based on the logarithm of the sample variance of the replicated observations at each design point. An important advantage of dealing with replicated observations in this context is that incorrect identification of the location effects has no impact on the dispersion model.

As it is much more difficult to identify significant changes in the spread than in the location, authors like Shoemaker, Tsui and Wu (1991), Bérubé and Nair (1998), Steinberg and Burzstyn (1998) and Myers, Khuri and Vining (1992) have shown that it is advan-

tageous to identify the most important sources of variation prior to the experiment and vary them systematically during the experiment. Estimating a response surface model in which the control by noise factor interactions are included is then an efficient way of identifying location and dispersion factors. In most response surface models however, the residuals are assumed to be homoscedastic, which means that all the important noise factors are assumed to have been identified and put under the control of the experimenter. As this is not very realistic, Engel and Huele (1996) have generalized this model to deal with heteroscedastic response surface models. Their iterative procedure is able to estimate the model once the significant location and dispersion effects have been identified but they do not deal with the model selection issue.

To model the residual variance based on just one observation at each design point is a challenging problem for several reasons. It requires the detection of significant effects in a saturated model which is always complicated, see Hamada and Balakrishnan (1998) for an overview of various approaches suggested to analyze unreplicated location models. The identification of significant dispersion effects is even more cumbersome because the only information about the dispersion is very imprecise as it is retrieved from the residuals of the location model. It is obvious that the dispersion model is very vulnerable to misspecification of the active location factors. But even when the correct location factors are identified, testing for significance of the dispersion effects is complicated due to the inefficiency of the testing procedures, partly because the location effects are estimated imperfectly. Especially when the dispersion effect to be tested is not present in the location model, most procedures that have been suggested perform poorly, see for instance Brenneman and Nair (2001) and McGrath and Lin (2001) for a comparison of the available test procedures. These authors also show that most of the procedures perform badly when there is more than one active dispersion factor and that they often suffer from spurious interaction effects such that two active dispersion effects may lead to incorrect detection of their interaction effect as significant too.

We will compare our method for model selection with the test procedures of Bergman

and Hynén (1997) and McGrath and Lin (2001) that have been shown to perform best for loglinear dispersion models. We refer to these procedures as B&H and M&L in the sequel. They are both based on the residuals \tilde{r}_i of an expanded location model including the active location effects, the factor k whose dispersion effect is tested for significance and all interactions between significant location factors and the factor k . So for each dispersion factor the corresponding expanded location model has to be estimated. The test statistics are

$$D_k^{B\&H} = \frac{\sum_{k+} \tilde{r}_j^2}{\sum_{k-} \tilde{r}_j^2} \quad (1)$$

$$D_k^{M\&L} = \frac{\prod_{k+} \tilde{r}_j^2}{\prod_{k-} \tilde{r}_j^2} \quad (2)$$

where \sum_{k+} and \sum_{k-} stand for the sum over the residuals at the $+$ and $-$ level of k respectively. It has been shown that the test statistics have approximate F-distributions, see McGrath and Lin (2001) for the distributions of $D_k^{B\&H}$ and $D_k^{M\&L}$. The test B&H has been shown to perform well when there is only one active dispersion effect. The method of M&L can cope with several active dispersion effects and does not suffer from spurious interaction detection but it is less powerful.

No matter how the active location and dispersion effects have been identified, once the model selection has been performed, an iterative procedure is typically used to get the final parameter estimates. REML (restricted maximum likelihood), IWLS (iteratively weighted least squares) or PML (pseudo maximum likelihood) have been suggested in the literature, see Engel and Huele (1996), Nelder and Lee (1991), Myers, Khuri and Vining (1992), Engel (1992), Grego (1993), and Lee and Nelder (2003). Wolfinger and Tobias (1998) use REML to fit a mixed model including location and dispersion effects as well as random effects.

In this paper we introduce a hierarchical Bayesian model in which the location and dispersion model are estimated simultaneously and select the significant effects based on the posterior information obtained. As the number of observations is typically small, the posterior information depends quite heavily on the prior information specified. We

propose an empirical Bayes method to specify the prior information based on the data and show through a simulation study that the posterior information obtained by this procedure is reliable to be used for the identification of the active location and dispersion effects.

In the next section we introduce the hierarchical model that we use to analyze the data. The Bayesian estimation procedure is explained in section 3. Section 4 deals with the empirical Bayes choice of the prior information which is crucial if the posterior information is to be used for model selection. Simulation studies are used in section 6 to illustrate the approach and to show that the performance is at least as good as the existing procedures. In section 7 we apply our method to data from an injection molding experiment and we end with a conclusion in section 8.

2 The Hierarchical Bayesian Model

Assume that the n responses y_i are modeled by the following heteroscedastic model:

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \sigma_i\epsilon_i \quad (i = 1, \dots, n) \quad (3)$$

with \mathbf{x}_i the p -dimensional vector with the coded factor level combinations of the potential location factors at run i and $\boldsymbol{\beta}$ the p -dimensional vector of the unknown location parameters of interest. The residuals ϵ_i are assumed to be independent standard normally distributed random variables and the variances of the response variables are assumed to follow a multiplicative model:

$$\log(\sigma_i^2) = \boldsymbol{\phi}'\mathbf{z}_i + w_i \quad (i = 1, \dots, n) \quad (4)$$

where the \mathbf{z}_i are q -dimensional vectors consisting of the coded factor levels of the potential dispersion factors at run i . Often, \mathbf{z}_i will contain a subset of the effects in \mathbf{x}_i . As the vectors \mathbf{x}_i and \mathbf{z}_i can also contain the noise factors that are kept under control during the experiment, the model can also be used to analyze a heteroscedastic response surface model. Finally, the residuals w_i are assumed to be independent and

normally distributed with mean zero and variance κ^2 . The unknown dispersion effects are contained in the q -dimensional vector $\boldsymbol{\phi}$.

The use of Bayesian methods requires prior distributions on the unknown parameters to be specified. As prior distributions for the parameters of interest $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, we will use the following p - and q -variate normal distributions:

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \Sigma_{\boldsymbol{\beta}}) \quad (5)$$

and

$$\boldsymbol{\phi} \sim N_q(\boldsymbol{\phi}_0, \Sigma_{\boldsymbol{\phi}}). \quad (6)$$

In section 4 we will provide an empirical Bayes method for choosing the values of the hyperparameters $\boldsymbol{\beta}_0$, $\Sigma_{\boldsymbol{\beta}}$, $\boldsymbol{\phi}_0$, $\Sigma_{\boldsymbol{\phi}}$ and κ^2 .

3 Model Estimation

In this section, it is explained how Gibbs sampling can be used for fitting the hierarchical Bayesian model outlined in the previous section. As in all Bayesian statistical procedures, the prior distribution of the unknown parameters which expresses the uncertainty about the parameters is updated by the information in the data to obtain the posterior distribution of the parameters. The averages of the posterior distributions can be used as point estimates for the parameters. The advantage of this approach is that the posterior distribution contains all the relevant information for any sample size, whereas the classical approach requires the researcher to rely on asymptotic formulas for the sampling distribution that need not be meaningful for the small samples that are typically used in experimental designs and we avoid some problems that are encountered in unreplicated designs.

The estimates for $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ and $\sigma_1^2, \dots, \sigma_n^2$ are obtained using Gibbs sampling combined with the Metropolis-Hastings algorithm, two related computationally intensive techniques which involve iterative resampling of all the model parameters, see Gelman, Carlin,

Stern and Rubin (2003). Let the n -dimensional vector \mathbf{y} be the vector containing the observations $y_i (i = 1, \dots, n)$. Given the structure of the model it is natural to implement the Gibbs sampler as follows

- initialize the chain at $\boldsymbol{\beta}^{(0)}, \boldsymbol{\phi}^{(0)}, \sigma_1^{2(0)}, \dots, \sigma_n^{2(0)}$
- at iteration k :
 - sample $\boldsymbol{\beta}^{(k)}$ from $p(\boldsymbol{\beta} | \boldsymbol{\phi}^{(k-1)}, \sigma_1^{2(k-1)}, \dots, \sigma_n^{2(k-1)}, \mathbf{y})$,
 - sample $\boldsymbol{\phi}^{(k)}$ from $p(\boldsymbol{\phi} | \boldsymbol{\beta}^{(k)}, \sigma_1^{2(k-1)}, \dots, \sigma_n^{2(k-1)}, \mathbf{y})$,
 - sample $\sigma_i^{2(k)}$ from $p(\sigma_i^2 | \boldsymbol{\beta}^{(k)}, \boldsymbol{\phi}^{(k)}, \sigma_1^{2(k)}, \dots, \sigma_{i-1}^{2(k)}, \sigma_{i+1}^{2(k-1)}, \dots, \sigma_m^{2(k-1)}, \mathbf{y})$, $i = 1, \dots, n$.

Using the prior distributions in (5) and (6), the conditional posterior distributions of the model parameters can be derived from the full posterior distribution of the unknown model parameters (see the appendix for more details about the derivation).

Denoting by \mathbf{X} the $(n \times p)$ -dimensional design matrix with \mathbf{x}_i' as its i^{th} row, and by $\Sigma_{\boldsymbol{\epsilon}} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ the variance-covariance matrix of the responses, we obtain as the conditional posterior distribution for $\boldsymbol{\beta}$ the p -dimensional normal distribution with mean

$$(\mathbf{X}'\Sigma_{\boldsymbol{\epsilon}}^{-1}\mathbf{X} + \Sigma_{\boldsymbol{\beta}}^{-1})^{-1} (\mathbf{X}'\Sigma_{\boldsymbol{\epsilon}}^{-1}\mathbf{y} + \Sigma_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0) \quad (7)$$

and covariance matrix

$$(\mathbf{X}'\Sigma_{\boldsymbol{\epsilon}}^{-1}\mathbf{X} + \Sigma_{\boldsymbol{\beta}}^{-1})^{-1}. \quad (8)$$

It is obvious from these formulas that the estimator for $\boldsymbol{\beta}$ boils down to the weighted least squares estimator if $\Sigma_{\boldsymbol{\beta}}^{-1} = \mathbf{0}$ which means that a noninformative prior for $\boldsymbol{\beta}$ is used. In case an informative prior is used, one can obtain estimates for $\boldsymbol{\beta}$ even if $\mathbf{X}'\Sigma_{\boldsymbol{\epsilon}}^{-1}\mathbf{X}$ is singular.

Similarly, denoting by \mathbf{Z} the $(n \times q)$ -dimensional design matrix with \mathbf{z}_i' as its i^{th} row and by \mathbf{L} the vector $\mathbf{L}' = (\log \sigma_1^2, \log \sigma_2^2, \dots, \log \sigma_n^2)$, then the conditional posterior

distribution of ϕ is shown to be a q -dimensional normal distribution with mean

$$\left(\frac{1}{\kappa^2}\mathbf{Z}'\mathbf{Z} + \Sigma_{\phi}^{-1}\right)^{-1} \left(\frac{1}{\kappa^2}\mathbf{Z}'\mathbf{L} + \Sigma_{\phi}^{-1}\phi_0\right) \quad (9)$$

and covariance matrix

$$\left(\frac{1}{\kappa^2}\mathbf{Z}'\mathbf{Z} + \Sigma_{\phi}^{-1}\right)^{-1}. \quad (10)$$

Here an uninformative prior $\Sigma_{\phi}^{-1} = \mathbf{0}$ leads to the ordinary least squares estimates for ϕ and with informative priors one can get estimates no matter whether $\mathbf{Z}'\mathbf{Z}$ is invertible or not.

Unfortunately, the conditional posterior distribution of each of the σ_i^2 ($i = 1, \dots, n$) does not follow a standard form. In order to get a closed form for all posterior distributions, we have investigated several model specifications but we were not successful. Therefore, a Metropolis-Hastings step is needed to sample values for σ_i^2 from their conditional posterior distributions. Let $\tau_i = \log(\sigma_i^2)$ and denote by $\pi(\tau_i)$ the exact conditional posterior distribution of τ_i . A new candidate value of τ_i , τ_i^{new} is then generated from a density that approximates the exact conditional posterior $\pi(\tau_i)$; it is shown in the appendix that the following normal distribution can be used to approximate the exact posterior distribution:

$$f(\tau_i) \sim N\left(\frac{\frac{\log(y_i - \beta' \mathbf{x}_i)^2}{2} + \frac{\phi' \mathbf{z}_i}{\kappa^2}}{\frac{1}{2} + \frac{1}{\kappa^2}}, \frac{1}{\frac{1}{2} + \frac{1}{\kappa^2}}\right). \quad (11)$$

The new sampled value is accepted as the new value for τ_i with probability

$$\min\left(1, \frac{\pi(\tau_i^{new})/f(\tau_i^{new})}{\pi(\tau_i^{old})/f(\tau_i^{old})}\right). \quad (12)$$

Once the sampling distributions have been derived, the Gibbs sampling and the Metropolis-Hastings algorithm can be implemented. It is however convenient to use the software package WinBUGS which can be downloaded free of charge from <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>. This software has been developed by the Biostatistics Unit in Cambridge, UK for the Bayesian analysis of complex statistical models

using Markov Chain Monte Carlo (MCMC) methods. The user only has to specify the model, the prior distributions and the initial values. Based on this information the software derives the conditional posterior distributions to be used in the Gibbs sampling and, if needed, even the approximating distribution to be used in the Metropolis-Hastings algorithm. The software can be used interactively or in batch-mode and several options are available to control the simulation process and to specify what statistics about the parameters and the model have to be displayed in the output. The WinBUGS software, version 1.4.1, was used to obtain the results in the sections 6 and 7. The necessary input files for WinBUGS as well as a SAS-program to create the necessary input files and to submit them to WinBUGS can be obtained from the authors.

4 Empirical Bayes Choice of Hyperparameters

As we deal with small sample sizes, the most critical point is the choice of the priors. If the choice of the priors is not at all informative, the posterior distributions of the parameters still have a very large variance, especially for the dispersion effects. In case the uncertainty in the prior information is very small, the data will not be able to overrule the information and the posterior mean will not deviate much from the prior mean. A similar dependence has been illustrated in Chipman, Hamada and Wu (1997) who analyze a homoscedastic location model with a Bayesian approach.

To get useful posterior information we use empirical Bayes methods in which the priors are specified based on the data. For the location parameters β it is quite intuitive to center the prior distribution of the intercept β_0 around the average response \bar{y} and of the other location parameters around zero. As best guess for the uncertainty of these prior distributions, we use the variance of the OLS estimates of the homoscedastic location model (or half the effects if the model is supersaturated). As the coding of the factors is arbitrary, the sign of the parameters is unimportant and we take the positive

and negative absolute values of the estimates into account:

$$\beta_0 \sim N(\bar{y}, \sigma_\beta^2); \quad \beta_{i>0} \sim N(0, \sigma_\beta^2) \quad (13)$$

$$\sigma_\beta^2 = \text{variance of } \{ \pm |\beta_{i>0}^{OLS}| \}; \quad \beta_i^{OLS} \text{ from } y_i = \beta' \mathbf{x}_i + \sigma \epsilon_i. \quad (14)$$

The prior distributions of the dispersion parameters are harder to specify as the final location model is unknown and this prior information is more critical for the reliability of the results. To avoid any assumption about the final location model, we use an empty location model to get some preliminary information about the dispersion parameters. As before, we center the independent normal prior distributions for the dispersion parameters around zero and use the variance of the OLS parameters as the corresponding degree of uncertainty:

$$\phi_0 \sim N(0, \sigma_\phi^2); \quad \phi_{i>0} \sim N(0, \sigma_\phi^2) \quad (15)$$

$$\sigma_\phi^2 = \text{variance of } \{ \pm |\phi_{i>0}^{OLS}| \}; \quad \phi_i^{OLS} \text{ from } \log(y_i - \bar{y})^2 = \phi' \mathbf{z}_i + w_i. \quad (16)$$

The only remaining hyperparameter left to be specified is κ^2 , the variance of the residual term in the dispersion model (4). We can use the general belief that after a model is fit to the data, the residual standard deviation will be roughly 20% of the unadjusted standard deviation in the response (see Chipman, Hamada and Wu (1997)). As we are dealing with a simplified dispersion model based on an empty location model in (16), 20% of the unadjusted standard deviation of $\log(y_i - \bar{y})^2$ will overestimate the final standard deviation of $\log \sigma_i^2$ considerably in most cases. Therefore we use 10% of the unadjusted standard deviation in the response $\log(y_i - \bar{y})^2$ as our best guess for the residual standard deviation κ . We show in the subsequent sections that this choice works quite well. Simulations have shown that putting a uniform prior on κ between 10% and 20% of the unadjusted standard deviation in the response $\log(y_i - \bar{y})^2$ leads to worse results, probably because of the extra uncertainty that is introduced in this case. The method proposed might however still benefit from a more sophisticated determination of κ which is a topic of future research.

As for the initialization step, simulations have shown that the results are independent of the initial parameter values for all reasonable numbers of iterations.

5 Model Selection

Model selection is based on the posterior distribution of the parameters. To identify active parameters, we use the credible or posterior intervals of the parameters; these are defined by the $\alpha/2$ th and $(1 - \alpha/2)$ th quantile of the posterior distribution. If the 95% credible interval of a parameter does not contain zero, this suggests that the parameter is significantly different from zero. Remark that these intervals have a very different interpretation than confidence intervals. As finding active dispersion effects is difficult it is best to defer the decision about the dispersion factors until the true location model has been identified, see McGrath and Lin (2001). Therefore we suggest the following procedure. First the model with the complete location and complete dispersion model is fitted and the significant location effects are identified based on the credible intervals. Secondly, the reduced location model and the complete dispersion model are estimated. We use these results to detect the significant dispersion effects. Finally, it is good practice to get final estimates by estimating the final location and dispersion model.

6 Simulation Results

Dispersion models

To start with, we present a simulation study without location effects to illustrate that the dispersion effects can be detected without too many false positives. We simulated 1000 datasets with 16 observations based on a 2^4 design from the model:

$$y_i = 1 + \sigma_i \epsilon_i \quad \text{with} \quad \log(\sigma_i^2) = \boldsymbol{\phi}' \mathbf{z}_i \quad (17)$$

and various active dispersion effects. In all simulations, we used the first 20000 iterations to achieve convergence and report the results about the posterior distributions based on the following 15000 iterations.

Table 1 shows the results for the dispersion models I , $I + 2A$, $I + 2A + 2B$, $I + 2A +$

$2B + 2AB$, $I + 2A + 2B + 2C$. The percentage of times that an effect has been labeled active using the 95% credible intervals together with the average mean of the posterior distribution are presented.

From the results for all these dispersion models it can be derived that the method proposed is able to identify the active dispersion effects quite well without making many type I errors. The results for dispersion models $I + 2A + 2B$ and $I + 2A + 2B + 2C$ show that the method does not suffer heavily from the spurious interaction bias while the results for $I + 2A + 2B + 2AB$ show that interaction terms are identified when active.

It is clear from the average estimates that the effects are all underestimated which is a known phenomenon in the estimation of dispersion models, see Brenneman and Nair (2001). At the bottom of the table we included the average final parameter estimates of the reduced dispersion model and the corresponding standard deviation over all simulations. This shows that the bias becomes smaller when the reduced dispersion model is fitted but it remains considerable.

Location and Dispersion effects

Here we repeat partly the simulation study of McGrath and Lin (2001) to be able to compare our results with the results obtained by the test statistics $D_k^{B\&H}$ and $D_k^{M\&L}$ in (1) and (2). We simulate 1000 datasets from a 2^4 design using the model:

$$y_i = 1 + A + B + C + AB + AC + BC + ABC + \sigma_i \epsilon_i \quad (18)$$

and the following dispersion models:

$$\log(\sigma_i^2) = I + 1.6A, \quad (19)$$

$$\log(\sigma_i^2) = I + 1.6A + 1.1C, \quad (20)$$

$$\log(\sigma_i^2) = I + 1.6D, \quad (21)$$

$$\log(\sigma_i^2) = I + 1.6D + 1.1BD. \quad (22)$$

Table 1: The percentage of times a dispersion effect has been declared significant using 95% credible intervals, and the average mean of the posterior distribution.

effect	active dispersion effects, parameters = 2									
	none		A		A, B		A, B, AB		A, B, C	
I	13	0.27	32	0.49	47	0.64	60	0.66	69	0.81
A	3	-0.09	88	1.24	77	1.12	96	1.33	78	1.02
AB	3	0.03	2	0.04	13	0.35	99	1.46	13	0.33
ABC	3	0.00	4	0.01	3	0.01	3	-0.00	5	-0.17
ABCD	2	0.01	1	0.02	2	0.02	1	-0.02	2	0.03
ABD	2	-0.00	2	-0.02	2	-0.02	1	-0.01	1	-0.01
AC	2	0.03	3	0.06	4	0.07	3	0.07	18	0.40
ACD	4	-0.00	3	-0.05	2	-0.07	1	-0.03	3	-0.04
AD	3	-0.01	4	-0.04	3	-0.00	2	0.01	5	-0.02
B	2	0.01	2	0.02	86	1.20	98	1.45	92	1.14
BC	3	-0.01	3	-0.02	4	0.02	2	-0.00	18	0.32
BCD	3	-0.02	2	-0.02	3	-0.06	3	-0.05	4	-0.02
BD	1	-0.04	1	-0.06	2	-0.05	1	-0.04	2	-0.04
C	3	-0.07	3	-0.10	2	-0.07	3	-0.09	87	1.06
CD	3	0.13	4	0.19	5	0.17	6	0.18	5	0.17
D	4	0.00	3	0.01	4	0.05	4	0.04	4	-0.01
		final $\hat{\phi}$		final $\hat{\phi}$		final $\hat{\phi}$		final $\hat{\phi}$		
		I 0.83		I 0.89		I 0.92		I 1.02		
		stdev 0.29		stdev 0.28		stdev 0.27		stdev 0.27		
		A 1.51		A 1.43		A 1.45		A 1.29		
		stdev 0.36		stdev 0.36		stdev 0.32		stdev 0.36		
				B 1.43		AB 1.52		B 1.40		
				stdev 0.35		stdev 0.30		stdev 0.39		
						B 1.60		C 1.39		
						stdev 0.32		stdev 0.36		

It is assumed that the correct location model has been identified and we include all terms up to the fourth order in the dispersion model to select the most significant ones. The procedures of B&H en M&L only allow to test dispersion effects that are also in the location model because the expanded location model is otherwise saturated. Tables 2 and 3 report the percentage of times that an effect is detected based on 95% credible intervals, as well as the results from the B&H en M&L tests with 5% significance level. As noted before B&H is able to detect sole dispersion effects quite well but at the expense of a large number of false positives. M&L keeps the required 5% type I errors

but has lower power, see also McGrath and Lin (2001) for a discussion of both tests. It is clear from table 2 that our procedure is capable of detecting the active dispersion effects if they are present in the location model with comparable power as M&L. From the left panel of table 3 it is quite clear that also this method does not pick up the dispersion effect D that is not in the location model. The right panel of table 3 shows the results for two dispersion effects D and BD that are not in the location model, but their interaction B is. Our method does pick up the two dispersion effects in 25% of the cases whereas B&H and M&L are not able to detect these effects at all.

Table 2: The percentage of times the effect has been declared significant using 95% credible intervals, and by the B&H and M&L tests at a 5% level.

	dispersion model $I + 1.6A$			dispersion model $I + 1.6A + 1.1C$		
effect	HB	BH	ML	HB	BH	ML
I	53			96		
A	53	82	53	60	76	53
AB	0	14	5	3	20	5
ABC	0	14	5	1	20	5
ABCD	0	-	-	3	-	-
ABD	0	-	-	3	-	-
AC	1	14	5	0	37	5
ACD	0	-	-	1	-	-
AD	0	-	-	4	-	-
B	4	14	5	6	20	5
BC	2	14	5	5	20	5
BCD	0	-	-	2	-	-
BD	0	-	-	7	-	-
C	2	14	5	41	48	26
CD	0	-	-	2	-	-
D	0	-	-	2	-	-

7 Numerical Example

The procedure is further illustrated by means of the injection molding experiment that was already described and analyzed by Engel (1992), Steinberg and Bursztyn (1994)

Table 3: The percentage of times the effect has been detected as significant using 95% credible intervals and by the B&H and M&L tests at a 5% level.

	dispersion model $I + 1.6D$			dispersion model $I + 1.6D + 1.1BD$		
	HB	BH	ML	HB	BH	ML
I	93			100		
A	6	5	5	6	9	5
AB	5	5	5	5	9	5
ABC	4	5	5	5	9	5
ABCD	0	-	-	1	-	-
ABD	0	-	-	0	-	-
AC	3	5	5	4	9	5
ACD	0	-	-	1	-	-
AD	2	-	-	4	-	-
B	6	5	5	25	33	13
BC	3	5	5	5	9	5
BCD	0	-	-	0	-	-
BD	8	-	-	25	-	-
C	11	5	5	6	9	5
CD	0	-	-	0	-	-
D	0	-	-	24	-	-

and Engel and Huele (1996). The goal of this experiment was to find the settings of the process parameters that resulted in a percentage of shrinkage close to a target value while keeping the amount of shrinkage insensitive to environmental factors such as the ambient temperature. In this study, several important noise factors were identified and included in the experiment in order to find significant noise by control factors. The experimental design used for varying the seven control factors and three noise factors was a Taguchi $L_8 \times L_4$ crossed array. The design and the data for the experiment are displayed in Table 4. In the table, the seven control factors are labelled A , B , C , D , E , F and G , and the three noise factors are called M , N and O . The table shows that the design for the control factors (the inner array of the product design) consists of eight test combinations and the design for the noise factors (the outer array) contains four combinations. Engel (1992) analyzed this experiment using a generalized linear model with nonconstant dispersion parameter and a power-of-the-mean variance function. Engel and Huele (1996) fitted a heteroscedastic hierarchical model through

iteratively reweighted least squares after selection of the important effects based on normal probability plots. Steinberg and Bursztyn (1994) considered the 32 observations as coming from a 2_{III}^{10-5} design and analyzed the results taking into account the alias structure of that highly fractionated resolution III design. They also use normal probability plots for model selection. We use their final models as benchmarks for the Bayesian approach advocated here.

Table 4: Design and data of injection molding experiment.

								M:	-1	-1	1	1
								N:	-1	1	-1	1
								O:	-1	1	1	-1
run	A	B	C	D	E	F	G	y ₁	y ₂	y ₃	y ₄	
1	-1	-1	-1	-1	-1	-1	-1	2.2	2.1	2.3	2.3	
2	-1	-1	-1	1	1	1	1	2.5	0.3	2.7	0.3	
3	-1	1	1	-1	-1	1	1	0.5	3.1	0.4	2.8	
4	-1	1	1	1	1	-1	-1	2.0	1.9	1.8	2.0	
5	1	-1	1	-1	1	-1	1	3.0	3.1	3.0	3.0	
6	1	-1	1	1	-1	1	-1	2.1	4.2	1.0	3.1	
7	1	1	-1	-1	1	1	-1	4.0	1.9	4.6	2.2	
8	1	1	-1	1	-1	-1	1	2.0	1.9	1.9	1.8	

A = cycle time, B = mould temperature, C = cavity thickness,

D = holding pressure, E = infection speed, F = holding time, G = gate size,

M = percentage regrind, N = moisture content, O = ambient temperature

The location model in (3) can be rewritten to make the distinction between control factors and noise factors more clear:

$$y_i = \beta'_x \mathbf{x}_i + \beta'_n \mathbf{n}_i + \mathbf{x}'_i \mathbf{\Lambda} \mathbf{n}_i + \sigma_i \epsilon_i \quad (23)$$

with

$$\log(\sigma_i^2) = \phi' \mathbf{z}_i + w_i \quad (24)$$

and \mathbf{n}_i the vector containing the factor level combination at run i of the noise factors that are controlled during the experiment. As in Engel and Huele (1996), the location effects tested are the ten main effects of the control and noise factors as well as all second order interactions between control and noise factors while the dispersion model contains the seven main effects of the control factors.

Table 5 contains the estimates of the complete location and dispersion model and \tilde{p} , twice the probability that the parameter is negative, in case the posterior mean is positive, and vice versa. The information in \tilde{p} can be used similarly as standard p -values instead of using the credible intervals as we did in the simulation study.

Table 5: The parameter estimates and \tilde{p} values for the full location and full dispersion model.

effect	$\hat{\beta}$	\tilde{p}	effect	$\hat{\beta}$	\tilde{p}	effect	$\hat{\phi}$	\tilde{p}
I	2.24	0.00	DO	-0.01	0.90	I	-1.12	0.02
A	0.27	0.02	E	0.20	0.20	A	0.01	0.98
AM	-0.04	0.72	EM	0.14	0.25	B	0.00	0.98
AN	-0.01	0.94	EN	-0.35	0.02	C	0.15	0.78
AO	0.01	0.94	EO	0.00	0.98	D	-0.10	0.78
B	-0.06	0.58	F	-0.00	0.98	E	0.12	0.82
BM	0.04	0.66	FM	-0.03	0.76	F	0.35	0.48
BN	0.01	0.96	FN	0.01	0.96	G	0.13	0.80
BO	-0.00	0.98	FO	0.02	0.84			
C	0.05	0.64	G	-0.14	0.22			
CM	-0.08	0.42	GM	0.01	0.90			
CN	0.37	0.00	GN	0.00	0.98			
CO	-0.01	0.94	GO	0.01	0.86			
D	-0.20	0.10	M	-0.03	0.72			
DM	-0.07	0.52	N	-0.00	0.98			
DN	-0.01	0.94	O	0.01	0.96			

It is clear from this table that effect A , CN and EN are active in the location model. Factor D is a borderline case. Therefore we fitted both reduced location models, with and without D and came up with the same active dispersion effect. We only report the results for the location model with D here. In table 6 we present the results of the reduced location and full dispersion model. The location parameters all remain significant and only the dispersion effect F seems to have a significant effect on the residual variance. Table 7 gives an overview of the location and dispersion effects that have been identified in previous studies. The significant control by noise interactions CN and EN were also discovered by Steinberg and Bursztyn (1994) and Engel and Huele (1996). Unlike them, we do not find evidence that G is an active location factor. Similar to Engel (1994), we find that F is the only active dispersion factor to model the residual variance.

Table 6: The parameter estimates and \tilde{p} values for the reduced location and full dispersion model.

effect	$\hat{\beta}$	\tilde{p}
I	2.30	0.00
A	0.32	0.02
CN	0.46	0.00
D	-0.28	0.00
EN	-0.44	0.00
effect	$\hat{\phi}$	ppv
I	-1.71	0.00
A	0.22	0.54
B	0.38	0.30
C	0.14	0.76
D	0.22	0.54
E	-0.33	0.36
F	0.73	0.05
G	0.06	0.90

8 Conclusion

A hierarchical Bayesian model was introduced to estimate and analyze the location and the dispersion model in a robust parameter design experiment. An empirical Bayes approach to choose the hyperparameters was presented. This choice is based on the ordinary least squares estimates of the corresponding homoscedastic location model and of a simplified dispersion model. Through simulation studies and a real life example it was shown that the resulting posterior information is reliable for the identification of the active location and dispersion factors. The method yields results that are competitive with other model selection methods suggested in the literature.

Table 7: An overview of models selected in the literature for the injection molding data.

method	location effects	dispersion effects
GLM, nonconstant dispersion factor, variance as power of the mean Engel (1992)	I, A, D, E	I, F
homoscedastic response surface model Steinberg & Bursztyn (1994)	I, A, D, G, CN, EN	I
heteroscedastic response surface model Engel & Huele (1996)	I, A, D, G, CN, EN	I, A
hierarchical Bayesian approach	I, A, D, CN, EN	I, F

Appendix

The derivation of the conditional posterior distributions of the different parameters in the hierarchical model starts from the full conditional posterior distribution of the parameters:

$$\begin{aligned}
& p(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma_1^2, \dots, \sigma_m^2 | \mathbf{y}) \\
& \propto p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\phi}) \left(\prod_{i=1}^n p(\sigma_i^2 | \boldsymbol{\phi}) \right) p(\boldsymbol{\beta}) p(\boldsymbol{\phi}) \\
& \propto \frac{1}{|\boldsymbol{\Sigma}_\epsilon|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) \\
& \quad \times \prod_{i=1}^n \frac{1}{\sigma_i^2} \exp \left(-\frac{(\log(\sigma_i^2) - \boldsymbol{\phi}' \mathbf{z})^2}{2\kappa^2} \right) \\
& \quad \times \exp \left(-\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right) \\
& \quad \times \exp \left(-\frac{1}{2} (\boldsymbol{\phi} - \boldsymbol{\phi}_0)' \boldsymbol{\Sigma}_\phi^{-1} (\boldsymbol{\phi} - \boldsymbol{\phi}_0) \right)
\end{aligned}$$

Posterior distribution for $\boldsymbol{\beta}$

Retrieving all terms including $\boldsymbol{\beta}$ in this expression leads to the following conditional posterior distribution of $\boldsymbol{\beta}$:

$$\begin{aligned}
p(\boldsymbol{\beta} | \boldsymbol{\phi}, \sigma_1^2, \dots, \sigma_m^2, \mathbf{y}) & \propto \tag{A.1} \\
& \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) \exp \left(-\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right).
\end{aligned}$$

The righthandside of this expression is proportional to

$$\exp \left(-\frac{1}{2} (\boldsymbol{\beta} - \mathbf{V}_\beta [\mathbf{X}\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{y} + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_0])' \mathbf{V}_\beta^{-1} (\boldsymbol{\beta} - \mathbf{V}_\beta [\mathbf{X}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{y} + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_0]) \right)$$

where

$$\mathbf{V}_\beta = (\mathbf{X}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{X} + \boldsymbol{\Sigma}_\beta^{-1})^{-1}. \quad (\text{A.2})$$

Posterior distribution for $\boldsymbol{\phi}$

In a similar fashion, the conditional posterior distribution for $\boldsymbol{\phi}$ can be obtained:

$$p(\boldsymbol{\phi}|\boldsymbol{\beta}, \sigma_1^2, \dots, \sigma_m^2, \mathbf{y}) \propto \left[\prod_{i=1}^n \exp \left(-\frac{(\log \sigma_i^2 - \boldsymbol{\phi}'\mathbf{z}_i)^2}{2\kappa^2} \right) \right] \exp \left(-\frac{1}{2} (\boldsymbol{\phi} - \boldsymbol{\phi}_0)' \boldsymbol{\Sigma}_\phi^{-1} (\boldsymbol{\phi} - \boldsymbol{\phi}_0) \right). \quad (\text{A.3})$$

The righthandside of this expression is proportional to

$$\exp \left(-\frac{1}{2} \left(\boldsymbol{\phi} - \mathbf{V}_\phi \left[\frac{1}{\kappa^2} \mathbf{Z}'\mathbf{L} + \boldsymbol{\Sigma}_\phi^{-1}\boldsymbol{\phi}_0 \right] \right)' \mathbf{V}_\phi^{-1} \left(\boldsymbol{\phi} - \mathbf{V}_\phi \left[\frac{1}{\kappa^2} \mathbf{Z}'\mathbf{L} + \boldsymbol{\Sigma}_\phi^{-1}\boldsymbol{\phi}_0 \right] \right) \right),$$

where

$$\mathbf{V}_\phi = \left(\frac{1}{\kappa^2} \mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}_\phi^{-1} \right)^{-1}. \quad (\text{A.4})$$

Posterior distribution for the variances

Finally, the conditional posterior distributions for the variances σ_i^2 (i, \dots, n) can be written as

$$p(\boldsymbol{\Sigma}_i^2|\boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{y}) \propto \left[\frac{1}{\sigma_i} \exp \left(-\frac{(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2}{2\sigma_i^2} \right) \right] \left[\frac{1}{\sigma_i^2} \exp \left(-\frac{(\log \sigma_i^2 - \boldsymbol{\phi}'\mathbf{z}_i)^2}{2\kappa^2} \right) \right]. \quad (\text{A.5})$$

Transforming to $\tau_i = \log \sigma_i^2$ gives

$$p(\tau_i|\boldsymbol{\beta}, \boldsymbol{\phi}, y_1, \dots, y_n) \propto \exp \left(-\frac{\tau_i}{2} - \frac{\exp(-\tau_i)}{2} (y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2 \right) \exp \left(-\frac{(\tau_i - \boldsymbol{\phi}'\mathbf{z}_i)^2}{2\kappa^2} \right). \quad (\text{A.6})$$

Unlike the original expressions for the conditional posterior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, this expression cannot be simplified to some known density. Therefore, the first term

of this conditional probability is approximated by a normal distribution with the same mode and curvature as the original function. The mode is given by

$$\log(y_i - \beta' \mathbf{x}_i)^2 \quad (\text{A.7})$$

and the curvature at the mode is given by $-1/2$. As the mode of a normal distribution with location parameter μ and scale parameter σ^2 equals μ and the curvature at the mode is $-1/\sigma^2$, the likelihood can be approximated by a normal distribution with $\mu = \log(y_i - \beta' \mathbf{x}_i)^2$ and $\sigma^2 = 2$.

The conditional posterior distribution of τ_i can then be approximated by

$$f(\tau_i) \propto \exp\left(-\frac{[\tau_i - \log(y_i - \beta' \mathbf{x}_i)^2]^2}{2 * 2}\right) \exp\left(-\frac{(\tau_i - \phi' \mathbf{z}_i)^2}{2\kappa^2}\right), \quad (\text{A.8})$$

which is a normal distribution with mean

$$\frac{\frac{\log(y_i - \beta' \mathbf{x}_i)^2}{2} + \frac{\phi' \mathbf{z}_i}{\kappa^2}}{\frac{1}{2} + \frac{1}{\kappa^2}} \quad (\text{A.9})$$

and variance

$$\left(\frac{1}{2} + \frac{1}{\kappa^2}\right)^{-1}. \quad (\text{A.10})$$

References

- Bergman, B., Hynén, A. (1997), "Dispersion Effects From Unreplicated Designs in the 2^{k-p} Series", *Technometrics*, 39, 2, 191-198.
- Bérubé, J., Nair, V.N. (1998), "Exploiting the Inherent Structure in Robust Parameter Design Experiments", *Statistica Sinica*, 8, 43-66.
- Brenneman, W.A., Nair, V.N., (2001), "Methods for Identifying Dispersion Effects in Unreplicated Factorial Experiments: A Critical Analysis and Proposed Strategies", *Technometrics*, 43, 4, 388-405.
- Chipman, H., Hamada, M., Wu, C.F.J., (1997), "A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing", *Technometrics*, 39, 4, 372-381.
- Engel, J. (1992), "Modelling Variation in Industrial Experiments," *Applied Statistics*, 41, 579-593.
- Engel, J., Huele, F. (1996), "A Generalized Linear Modeling Approach to Robust Design", *Technometrics*, 38, 365-373.

- Gelman, A., Carlin, J., Stern, H., Rubin, D. (2003) *Bayesian Data Analysis, second edition*, Chapman & Hall.
- Grego, J.M. (1993), "Generalized Linear Models and Process Variation," *Journal of Quality Technology*, 25, 288-295.
- Hamada, M., Balakrishnan, N., (1998), "Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals", *Statistics Sinica*, 8, 1-41.
- Lee, Y., Nelder, J. (2003), "Robust Design via Generalized Linear Models", *Journal of Quality Technology*, 35, 2-12.
- McGrath, R.N., Lin, D.K.J., (2001), "Testing Multiple Dispersion Effects in Unreplicated Fractional Factorial Designs", *Technometrics*, 43, 4, 406-414.
- Myers, R.H., Khuri, A.I., Vining, G.G. (1992), "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician*, 46, 131-139.
- Nair, V.N. (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 127-161.
- Nelder, J.A., Lee, Y. (1991), "Generalized Linear Models for the Analysis of Taguchi-Type Experiments," *Applied Stochastic Models and Data Analysis*, 7, 101-120.
- Shoemaker, A.C., Tsui, K.-L., Wu, C.F.J. (1991), "Economical Experimentation Methods for Robust Parameter Design," *Technometrics*, 33, 415-428.
- Steinberg, D., Bursztyn, D (1994), "Confounded Dispersion Effects in Robust Design Experiments with Noise factors," *Journal of Quality Technology*, 26, 12-20.
- Steinberg, D., Bursztyn, D (1998), "Noise factors, dispersion effects and robust design," *Statistica Sinica*, 8, 67-85.
- Taguchi, G. (1987), *System of Experimental Design* (Vol. 1 and 2), Dearborn, MI: American Supplier Institute.
- Vining, G.G., Myers, R.H. (1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, 22, 38-45.
- Welch, W. J., T.-K. Yu, S. M. Kang, J. Sacks (1990) "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, 22, 15-22.
- Wolfinger, R.D., Tobias, R.D. (1998), "Joint Estimation of Location, Dispersion, and Random Effects in Robust Design," *Technometrics*, 40, 1, 62-71.